

Deriving a neutral model of species abundance from fundamental mechanisms of population dynamics

F. HE†

Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada T6G 2H1

Summary

1. Ecologists have long sought to derive assembly rules of ecological communities from the fundamental processes of population dynamics, but this goal has remained elusive. Neutral theory has reinvigorated the search by showing that patterns of relative species abundance closely resembling those actually observed arise under the assumption that, to a first approximation, all species are demographically identical on a per capita basis.
2. Here a neutral model is proposed to incorporate all four fundamental processes of population dynamics: birth, death, immigration and emigration. This symmetric model demonstrates that patterns of relative species abundance are fully derivable from these basic processes of population dynamics.
3. The theory derived extends the concept of community by showing that a continuum exists between large-scale ('metacommunity') and small-scale ('local community') processes, eliminating the artificial distinction between the two made by the current neutral theory and by the theory of island biogeography.
4. The population-based species-abundance model describes very well the observed patterns of relative species abundance of tropical trees, breeding birds in the USA, aphids at Rothamsted, UK, and estuarine fishes in the north-east USA.
5. The study also notes that while species assemblages may be well described by the neutral processes of population dynamics, the inference of mechanisms from pattern fitting is not warranted because one-to-one relationships between generating mechanisms and community patterns usually do not exist, either in the neutral realm or in the niche world.

Key-words: Community assembly rules, local community, metacommunity, migration rates, neutral theory, population demographics, spatial scale, species diversity

Functional Ecology (2005) **19**, 187–193

Introduction

The search for the mechanisms underlying species-abundance distributions continues to attract much attention (e.g. Magurran & Henderson 2003; McGill 2003a; Sugihara *et al.* 2003; Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003; Etienne & Olf 2004; McKane, Alonso & Solé 2004). Neutral theory in ecology has reinvigorated this search by showing that patterns of diversity closely resembling those observed in nature can arise under the assumption that organisms in a community have identical demographics on a per capita basis (Hubbell 2001). Recent development of this theory has concentrated on describing relative species abundances of neutral communities at meta-community and local community scales (Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003; McKane

et al. 2004). A metacommunity is defined as the self-contained evolutionary biogeographical unit within which most member species originate, live and die. Here, speciation is the analogue of immigration in a local community. In contrast, a local community is subject to an exchange of migrants with the metacommunity where it is embedded or with other local communities via immigration and emigration. The species-abundance distribution of the local community can be described by a zero-sum multinomial distribution (Hubbell 2001; McKane *et al.* 2004) or other models (Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003), while the log-series distribution under point-mutation speciation is the best-known metacommunity pattern (Hubbell 2001; Volkov *et al.* 2003).

The conceptual distinction between the metacommunity and the local community has played a critical role in the development of the spatially implicit version of the neutral theory of biodiversity (Hubbell 2001). Recent neutral models (Vallade & Houchmandzadeh

2003; Volkov *et al.* 2003; McKane *et al.* 2004) make the same distinction, treating the metacommunity as a source of immigrants but otherwise as dynamically unaffected by the local community. Species in these models are not treated as fully neutral-symmetric, because their relative abundances in the local community are functions of their relative abundances in the metacommunity. In the absence of immigration, these local community models do not converge to the metacommunity models (e.g. the zero-sum multinomial local model does not converge to the log-series distribution, see McKane *et al.* 2004). This is perfectly valid under the assumptions of the classic island-mainland system, in which local species abundances on islands depend, through immigration, on fixed species abundances in the metacommunity. Therefore, these models may be more accurately called island models.

However, in reality there is no sharp line that separates the metacommunity and the local community. On actual landscapes, there is a continuum from local communities to the large-scale metacommunity, varying in the degree to which their dynamics are affected by migration. By contrast with the island-mainland system, on such continuous landscapes local communities are embedded within the metacommunity. A very general approach to model such a system is to assume a fully symmetric local community where species abundances are each affected by local births and deaths, as well as by immigration and emigration. Here, I show that a simple consideration of these four fundamental processes of population dynamics, i.e. birth, death, immigration and emigration, allows one to derive a fully symmetric neutral model for the distribution of relative species abundance in communities on any scale.

The model

Consider the basic continuous time, discrete state, Markov chain for the stochastic dynamics of population growth (Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003; McKane *et al.* 2004):

$$\frac{dp_{n,k}(t)}{dt} = p_{n+1,k}(t)d_{n+1,k} + p_{n-1,k}(t)b_{n-1,k} - p_{n,k}(t)(b_{n,k} + d_{n,k}) \quad \text{eqn 1}$$

where $p_{n,k}(t)$ is the probability that the k th species contains n individuals at time t , and $b_{n,k}$ and $d_{n,k}$ are the rates of birth and death, respectively, of the species with n individuals. Further, set the boundary condition $p_{-1,k}(t) = d_{0,k} = 0$.

Suppose the community defined by equation 1 has total number of species S . Then the species-abundance distribution for the community is:

$$\langle \phi_n \rangle = \sum_{k=1}^S p_{n,k} \quad \text{eqn 2}$$

where $\langle \phi_n \rangle$ is the average number of species that have n individuals (Volkov *et al.* 2003). It is easy to show

from equation 1 that at steady state (i.e. $t \rightarrow \infty$) equation 2 is a function of the birth and death rates (Volkov *et al.* 2003; McKane *et al.* 2004):

$$\langle \phi_n \rangle = Sp_{0,k} \prod_{i=0}^{n-1} \frac{b_{i,k}}{d_{i+1,k}} \quad \text{eqn 3}$$

Several metacommunity and local community models have been developed by assuming various forms of birth and death rates (Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003; McKane *et al.* 2004). In particular, Volkov *et al.* (2003) have shown that under linear birth and death (i.e. $b_{n,k} = b_k n$, $d_{n,k} = d_k n$, where b_k and d_k are instantaneous birth and death rates, respectively, for the k th species), equation 3 leads to the log-series distribution for relative species abundance in the metacommunity. In the metacommunity, new species are added only by speciation (Hubbell 2001). Note that the subscript k is ignored here because the neutral theory assumes that all individuals, regardless of species identity, experience the same probabilities of birth and death.

I now generalize the log-series metacommunity model by considering the linear birth and death rates plus immigration and emigration:

$$b_{n,k} = b_k n + \lambda_k \quad \text{and} \quad d_{n,k} = d_k n + \mu_k, \quad \text{eqn 4}$$

where λ_k is the immigration rate measured by the number of immigrants per unit time and μ_k is emigration rate. Note that although immigration and emigration rates defined in the above are constant for a given population, they are assumed to differ from species to species as denoted by the subscript k . This assumption can be relaxed to derive the neutral model.

Under the neutrality assumption the species identity k can be ignored, substituting the rates in equation 4 into (3) leads to

$$\langle \phi_n \rangle = Sp_0 \frac{(v + \lambda)(b + \lambda) \dots ((n - 1)b + \lambda)}{(d + \mu)(2d + \mu) \dots (nd + \mu)}$$

where v is the speciation rate defined by $b_0 = v + \lambda$. This implies that when a species becomes extinct (i.e. $n = 0$ in equation 4) there are two possible ways to replace it with a new species. One is through speciation at rate v , and the other is through immigration at rate λ . Hence, the model is no longer a metacommunity model where migration is not allowed, but a local community model that couples with the metacommunity. Simplification of the above equation leads to the species-abundance distribution:

$$\langle \phi_n \rangle = \theta \frac{\Gamma(n + \alpha)\Gamma(1 + \beta)}{\Gamma(\alpha)\Gamma(n + 1 + \beta)} x^n, \quad n = 1, 2, \dots \quad \text{eqn 5}$$

where $\alpha = \lambda/b$, $\beta = \mu/d$ and $x = b/d$. $\theta = Sp_0(v/\lambda + 1)$ is not a free parameter but a normalizing factor that makes the summation of equation 5 over n be the total number of species S of a community (for frequency

distribution) or 1 (for probability distribution). It has the form:

$$\theta = \frac{1 + \beta}{\alpha x} \frac{1}{F(1 + \alpha, 2 + \beta, x)} \quad \text{eqn 6}$$

where $F(1 + \alpha, 2 + \beta, x)$ is a standard hypergeometric function $\sum_{i=0}^{\infty} \frac{\Gamma(1 + \alpha + i)}{\Gamma(1 + \alpha)} \frac{\Gamma(2 + \beta)}{\Gamma(2 + \beta + i)} x^i$ (see Zillinger 2003, p. 36); many mathematical programs, such as Maple or Mathematica, have standard built-in commands for evaluating hypergeometric functions.

Equation 5 is a model for a local community because the force of migration is in operation, but it is easy to show that in the absence of migration (i.e. $\lambda = \mu = 0$, or $\alpha = \beta = 0$), equation 5 simply reduces to the log-series distribution for a metacommunity. Furthermore, if there is no emigration (i.e. $\mu = 0$, or $\beta = 0$) it is a truncated negative binomial distribution, $\langle \phi_n \rangle = \theta \frac{\Gamma(n + \alpha)}{\Gamma(\alpha)} \frac{x^n}{n!}$, which can be obtained through applying the same linear growth with immigration to equation 1 (Boswell & Patil 1970; Taylor & Karlin 1998). The truncated negative binomial distribution is little known to ecologists but its applications to species-abundance distributions have indeed been discussed in the literature (Pielou 1975; He & Legendre 2002).

Parameter x in model 5 has the same meaning as in Fisher's log-series distribution, equal to the ratio of the per capita birth rate to per capita death rate (Volkov *et al.* 2003). In addition, the generalized model has two more parameters (α and β), which measure the relative strength in the community of immigration *vs* birth and emigration *vs* death, respectively. Adding the parameters for migration provides deeper insight into the effects of the four basic demographic processes of birth, death, immigration and emigration on the

species-abundance distribution (Figs 1 and 2). High α (i.e. high immigration relative to birth rate) leads to left-skewed species-abundance distributions, suggesting that there are few rare species (or more abundant species) in such communities (Fig. 1). This result is consistent with the prediction of metapopulation theory (Hanski 1991) or the source-sink effect of diversity (Pulliam 1988) which propose that constant immigration would rescue rare species from extinction by increasing their abundance. The rescue effect is operating here in a neutral manner, i.e. immigration rate is independent of the abundance of individual species, as is clear from $b_n = bn + \lambda$ (although birth rate depends linearly on abundance). Contrary to the effect of immigration, emigration skews the species-abundance distribution to the right, resulting in an increasing proportion of rare species (Fig. 2).

Similar to Fisher's log-series distribution for the metacommunity, θ in model 5 is also a biodiversity parameter which is a function of α , β and x as defined by equation 6. Note that θ in Fisher's log-series distribution (α in Fisher's notation) equals $-\ln(1 - x)$, which can also be deduced from the hypergeometric function of equation 5 in the absence of migration (i.e. $\alpha = 0$ and $\beta = 0$).

Empirical fitting of the model

Model 5 is now fitted to four species assemblages: trees, birds, aphids and fishes.

1. Tropical tree abundance: these are tree census data from the famous 50-ha stem-mapping plot on the Barro Colorado Island (BCI) of Panama. The abundance of each tree species is the number of stems with diameter at breast height ≥ 10 cm. The data were published by Condit *et al.* (2002) and used by McGill (2003a) and Volkov *et al.* (2003) to test the neutral theory.
2. The North American Breeding Bird Survey (BBS) data: 100 sets (routes) of the BBS data (Robbins, Bystrak & Geissler 1986; Sauer, Hines & Fallon 2001) were used by McGill (2003a) to test the neutral theory. In his study each data set (route) was an average of the bird count over a 5-year period (1996–2000). Here the data from the first route (46°42' N, 66°93' W) of McGill's 100 routes was used. Instead of using the average of the 5-year bird count, here the total count over the 5-year period was used because model 5 requires integer data.
3. Rothamsted aphids: the Rothamsted trap is part of the EXploitation of Aphid Monitoring systems IN Europe (EXAMINE; <http://www.rothamsted.bbsrc.ac.uk>). The daily abundance of flying aphids has been recorded for up to 40 years using a network of suction traps, each with its aperture 12.2 m above the ground, throughout Europe (Taylor 1986; Woivod & Harrington 1994; Harrington *et al.* 2004). Data presented here are the annual totals (male + female) of each species in samples from the trap sited at

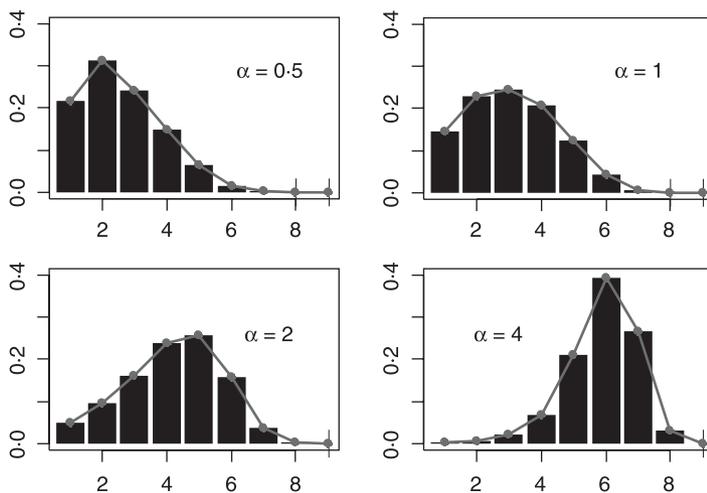


Fig. 1. Illustrating the effect of α (immigration relative to birth rate) on species abundance model 5 given $\beta = 1$, $x = 0.9$. High α skews the distribution to the left. The distributions were plotted using Preston's binning method as follows: the first bar is $\langle \phi_1 \rangle / 2$, the second bar is $\langle \phi_1 \rangle / 2 + \langle \phi_2 \rangle / 2$, the third bar is $\langle \phi_2 \rangle / 2 + \langle \phi_3 \rangle + \langle \phi_4 \rangle / 2$, the fourth bar is $\langle \phi_4 \rangle / 2 + \langle \phi_5 \rangle + \langle \phi_6 \rangle + \langle \phi_7 \rangle + \langle \phi_8 \rangle / 2$, etc. The numbers on the x -axis represent Preston's octave classes.

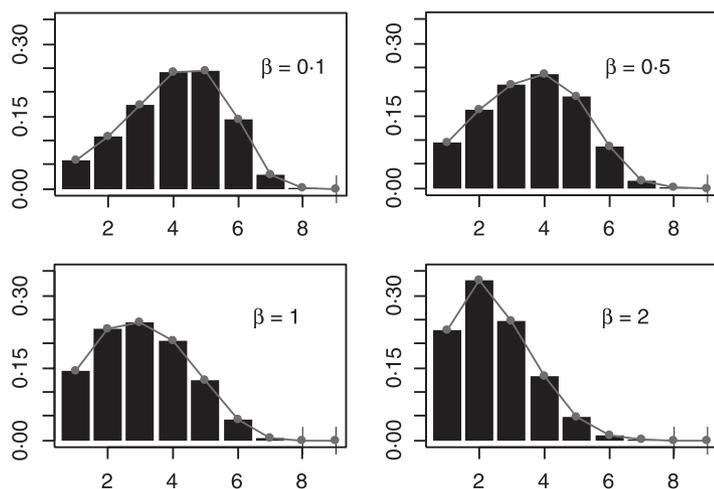


Fig. 2. Illustrating the effect of β (emigration relative to death rate) on species abundance model 5 given $\alpha = 1$, $x = 0.9$. High β skews the distribution to the right. The distributions were plotted using Preston's binning method as described in Fig. 1. The numbers on the x -axis represent Preston's octave classes.

Rothamsted Research, Harpenden, UK, in 2001. In the case where the number of days for a count is greater than one, the aphid count was divided by the number of days for the count and the result was rounded to the nearest integer value. This is the standard data reported by the EXAMINE system.

4. Estuarine fishes: these data were collected from 110 estuary stations located in the Virginian Provinces in the north-east USA (<http://www.epa.gov/emap/html/data/estuary/>) in July and August 1993. Fish abundance was enumerated in the field after completion of one successful standard trawl. The trawl net was a funnel-shaped, high-rise sampling trawl with a 16-m footrope with a chain sweep. The trawl net had 5-cm mesh wings and a 2.5-cm cod end.

The maximum likelihood function of model 5 is easy to compute. Given observed abundances of s species $\mathbf{n} = \{n_1, n_2, \dots, n_s\}$, the log-likelihood function of model 5 is $l(\alpha, \beta, x, \mathbf{n}) = \sum_{i=1}^S \log \langle \phi_n \rangle$. The maximum likelihood estimates of the three parameters (α, β, x) were obtained by maximizing this log-likelihood function. The maximization was evaluated using the iterative Newton–Raphson method using SPLUS.

Table 1. Maximum likelihood estimates (α, β and x), the Kolmogorov–Smirnov (KS) and χ^2 goodness-of-fit tests of model 5 to the four data sets of trees, birds, aphids and fishes. S is the total number of species, S_1 and S_2 are, respectively, the number of singletons and doubletons, N is the total number of individuals and θ is the biodiversity parameter from equation 6. In the χ^2 test, when the predicted number of species of the last bin in Fig. 3 is smaller than 1, the bin was merged with the bin to the left

Communities	Summary of the data			Parameters and tests of model 5					
	S	$S_1 + S_2$	N	θ	α	β	x	p_{KS}	p_{χ^2}
BCI trees	225	19 + 13	21 457	0.09622	2.8683	2.9654	0.9982	0.821	0.902
BBS birds	84	5 + 5	3 637	0.11047	1.0645	0.7053	0.9895	0.863	0.992
Rothamsted aphids	113	24 + 13	7 497	0.6833	0.6368	1.0256	0.9991	0.684	0.901
Estuarine fishes	523	168 + 77	13 007	621.544	0.0007699	0.4464	0.9988	0.944	0.972

Summaries of the four data sets and the results of model fitting are shown in Table 1. Both the χ^2 and the Kolmogorov–Smirnov tests indicate that model 5 describes the four communities very well. Because the χ^2 test is subject to different binning methods and is sensitive to the bins chosen, the Kolmogorov–Smirnov test is generally more powerful and should be relied on. Figure 3 shows the observed and predicted cumulative probability distributions of the species abundance (left-hand column) and their frequency distributions (right-hand column).

The log-normal (McGill 2003a) and the local model of Volkov *et al.* (2003) have been previously fitted to the data from BCI. Here, for the purpose of comparison, these two models are also fitted to the other three data sets and the results are shown in Table 2 and Fig. 3. The log-normal model describes all four data sets well. Volkov *et al.*'s model works well for the tropical tree and BBS bird data but fails to fit the aphid and fish data (Table 2 and Fig. 3). The common features of the aphid and fish data are that they have an excessive number of rare species and thus are highly skewed to the right. The failure of the Volkov *et al.* model for communities of great numbers of rare species is possibly because emigration is not considered in their model. It is interesting to observe that, while the Volkov *et al.* model fails the aphid and fish data, the fits of the log-normal model and model 5 to these data are very close. However, as has been repeatedly argued, the statistical goodness-of-fit should not be the ultimate criterion for judging a model, particularly when, as with a number of species-abundance distributions, the differences between models may be slight (McGill 2003b; Sugihara *et al.* 2003). The underlying biological mechanisms of the models are a much more important criterion (Ginzburg & Jensen 2004). Model 5 and the Volkov *et al.* model have clear advantages in this aspect. For application, however, the Volkov *et al.* model, which has a complex integration term, is difficult to solve quickly and accurately.

Discussion

The generalized model 5 is a promising addition to the family of neutral models of biodiversity because of its simplicity, its reduced number of assumptions and its

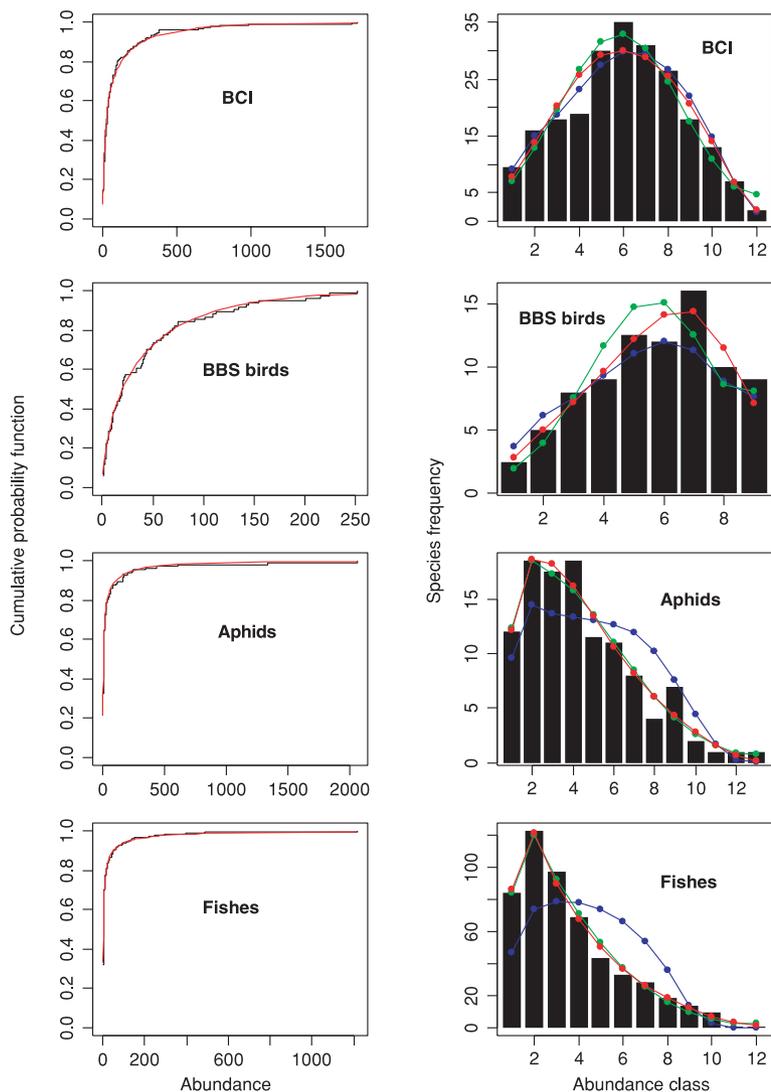


Fig. 3. Red curves are the fits of the generalized model 5 to the BCI trees, BBS birds, Rothamsted aphids and estuarine fishes (see Table 1). On the left-hand column are the cumulative probability functions on which the Kolmogorov–Smirnov test was based. The step curves are the observed cumulative probabilities, and the smooth red curves are the predictions. On the right-hand column are the species frequency distributions on which the χ^2 test was based. The fits of the log-normal distribution (green curves) and Volkov *et al.* model (blue curves) are also included in the frequency plots for comparison (see Table 2). The frequency distributions were plotted using Preston's binning method as follows: the first bar is $<\phi_1>/2$, the second bar is $<\phi_1>/2 + <\phi_2>/2$, the third bar is $<\phi_2>/2 + <\phi_3> + <\phi_4>/2$, the fourth bar is $<\phi_4>/2 + <\phi_5> + <\phi_6> + <\phi_7> + <\phi_8>/2$, etc.

generality. Similar to other neutral models (Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003; McKane *et al.* 2004), model 5 assumes that individuals are demographically identical. In addition to this assumption, current neutral theory requires that a steady-state community must meet two more conditions: the conservation of community size (the total number of individuals in a community) and the total number of species (Hubbell 2001; Vallade & Houchmandzadeh 2003; Volkov *et al.* 2003; McKane *et al.* 2004). To derive the log-series metacommunity model, total species number must be conserved (Volkov *et al.* 2003), and to derive the zero-sum multinomial model of

McKane *et al.* (2004) or the local community model of Volkov *et al.* (2003), community size must also be conserved. Model 5, which unifies both local and metacommunities, is derived from equation 2, and therefore requires only conservation of the total number of species, not the conservation of community size. The conservation of community size assumed by earlier neutral models has been criticized as too restrictive to be realistic. The relaxation of the conservation of community size is of practical significance to the development of more realistic community theories because this condition is unlikely to be met by many communities in reality. In addition, model 5 is mathematically simpler than the zero-sum multinomial distribution of Hubbell (2001) and Volkov *et al.* (2003) model. At equilibrium, although the total number of species in the generalized model 5 is preserved, species identity is not. Species may become extinct due to death or emigration. As soon as that occurs (i.e. the abundance of the species n hits 0), a new species will appear in a given community through speciation or immigration or both, as defined by $b_0 = v + \lambda$. Note that in the case of the metacommunity, $\lambda = 0$, and so speciation is the only force for the appearance of a new species.

The theory developed in this study has plausibly laid a population foundation for understanding macroecological patterns of species abundance in terms of the four fundamental processes controlling population dynamics: birth, death, immigration and emigration. It is however, important to note that while these processes are sufficient for deriving the generalized species abundance model 5, they are not a necessary condition. Besides the set of processes described by equation 4, it is easy to show that another set of processes that can lead to the same form of model 5 is density-dependent birth and death rates as defined by $b_n = b(n + a)n + \lambda$ and $d_n = d(n + c)n$. Similarly, the metacommunity log-series model can also be generated by density-dependent birth and death rates of $b_n = bn(n + 1)$ and $d_n = dn^2$, which differ from the linear birth and death rates $b_n = bn$ and $d_n = dn$ given by Volkov *et al.* (2003). The ambiguity in mechanisms in generating community patterns is not a unique problem to neutral models, but is a prevailing problem in community ecology in general (see Gaston & Blackburn 2000 for a thorough review). For example, the log-normal distribution could be equally well explained by the application of the large number theorem to the statistical process of multiplicative products (Pielou 1975) or by the assumption of the population growth of the Gumpertz form in a stochastic environment (Engen & Lande 1996). Another example is the species–area relationship whose genesis can be explained by many competing processes (McGuinness 1984). The non-unique relationship between mechanisms and patterns suggests that we should not infer mechanisms from patterns or do so only with extreme caution. The reason that the interpretation of model 5 by the linear birth + immigration and linear death + emigration is

Table 2. Fits of the log-normal distribution and the local model of Volkov *et al.* (2003) to the four data sets of trees, birds, aphids and fishes. The log-normal model, $\langle \phi_n \rangle = (N_0/n) \exp\{-[\log_2(n) - \log_2(n_0)]^2/2\sigma^2\}$, was parameterized by the maximum likelihood method. N_0 in the model is not a free parameter but a normalizing factor making the sum of $\langle \phi_n \rangle$ equal 1. Because the maximum likelihood of the Volkov *et al.* model was difficult to solve, the model (with parameters: immigration rate m and the diversity parameter θ) was evaluated by a global search to maximize the P -values of the tests. In the χ^2 test, when the predicted number of species of the last bin in Fig. 3 is less than 1, the bin was merged with the bin to the left. Note that both models are expressed as frequency distributions rather than probability mass functions, and the estimates of m and θ for BCI plot are directly from Volkov *et al.* (2003)

Communities	Log-normal model				Volkov <i>et al.</i> model				
	N_0	n_0	σ^2	P_{KS}	P_{χ^2}	m	θ	P_{KS}	P_{χ^2}
BCI trees	47.925	20.348	5.348	0.575	0.769	0.1	47.226	0.683	0.941
BBS birds	22.204	17.389	3.343	0.693	0.765	0.0745	21.608	0.594	0.790
Rothamsted aphids	25.008	1.651	10.986	0.754	0.924	0.99	19.421	0.0152	0.233
Estuarine fishes	203.025	0.156	13.486	0.830	0.704	0.64	117.297	0.000	0.000

adopted in this study is because the mechanisms seem to be the simplest and most parsimonious.

Another potential problem with model 5 is that although it is perfectly valid to assume constant rates of immigration (λ) and emigration (μ) for a given population, there may potentially be a problem in applying equation 4 across all species in a community. The problem is not with the immigration rate but with the emigration rate. It is reasonable to assume a constant immigration rate across all species because no competition is involved in structuring a neutral local community, thus immigrants to a local community should be independent of the species already in the community. However, it seems less reasonable to assume a constant emigration rate across species because the outgoing species are supposed to depend on their abundances in the local community: abundant species should have higher emigration rates than rare species. This suggests that a realistic formulation for the emigration rate, μ_k , in equation 4 is to retain subscript k to identify the difference in emigration rates among species. However, these species-specific emigration rates overparameterize the community model and do not lead to any form of analytical solution to equation 3. A reasonable interpretation for the constant immigration and emigration rates required for deriving the generalized species abundance model is that the rates are considered as averages, and in that sense model 5 is therefore a simplified mean-field model (McKane, Alonso & Solé 2000).

In summary, the species-abundance distribution model developed here is a population-based model that ties community patterns to the four fundamental processes controlling population dynamics. The model provides a more general theoretical approach for unifying the study of community assembly on arbitrary spatial scales. The key conclusion of the theory is that the species-abundance relationship can be explained by the parsimonious (yet complete) set of fundamental demographic processes underlying population dynamics. Moreover, having a population-dynamical theory of community assembly will help bridge the gap between the neutral and niche-based theories, because both now

depend on the same set of fundamental population processes. What remains to be explored is what happens when we incorporate interspecific demographic differences into the species-abundance model.

Acknowledgements

F.H. is a Canada Research Chair. Thanks to Stephen Hubbell, Luis Borda de Agua, Phil DeVries, Kevin Gaston, Richard Harrington, Xinsheng Hu, Russell Lande, Pierre Legendre, Brian McGill, Vojtech Novotny and two referees for sharpening this paper. McGill and Harrington generously provided the BBS bird and Rothamsted aphid data, respectively. F.H. is grateful to the numerous volunteers, the Center for Tropical Forest Science (the BCI data), the EXAMINE consortium (the aphid data), the Patuxent Wildlife Research Center (the BBS data) and the US Environmental Protection Agency through its Environmental Monitoring and Assessment Program (EMAP) (the estuarine fish data) for their contributions to the data used in this study. Although the data used in this article have been funded by various agencies, it has not been subjected to review of the agencies and no official endorsement should be inferred. This work is supported by the NSERC (Canada) and the Canadian Forest Service.

References

- Boswell, M.T. & Patil, G.P. (1970) Chance mechanisms generating the negative binomial distributions. *Random Counts in Models and Structures* (ed. G.P. Patil), pp. 3–22. Pennsylvania State University Press, University Park, PA.
- Condit, R., Pitman, N., Leigh, E.G. Jr, Chave, J., Terborgh, J., Foster, R.B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H.C., Losos, E. & Hubbell, S.P. (2002) Beta-diversity in tropical forest trees. *Science* **295**, 666–669.
- Engen, S. & Lande, R. (1996) Population dynamic models generating the lognormal species abundance distribution. *Mathematical Biosciences* **132**, 169–183.
- Etienne, R.S. & Olf, H. (2004) A novel genealogical approach to neutral biodiversity theory. *Ecology Letters* **7**, 170–175.
- Gaston, K.J. & Blackburn, T.M. (2000) *Pattern and Process in Macroecology*. Blackwell Science, Oxford.

- Ginzburg, L.R. & Jensen, X.J. (2004) Rules of thumb for judging ecological theories. *Trends in Ecology and Evolution* **19**, 121–126.
- Hanski, I. (1991) Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society* **42**, 17–38.
- Harrington, R., Verrier, P., Denholm, C., Hullé, M., Maurice, D., Bell, N., Knight, J., Rounsevell, M., Cocu, N., Barbagallo, S., Basky, Z., Coceano, P.-G., Derron, J., Katis, N., Lukášová, H., Marrkula, I., Mohar, J., Pickup, J., Rolot, J.-L., Ruzkowska, M., Schliephake, E., Seco-Fernandez, M.-V., Sigvald, R., Tsitsipis, J. & Ulber, B. (2004) 'EXAMINE' (EXploitation of Aphid Monitoring in Europe): An EU thematic network for the study of global change impacts on aphids. *Aphids in a New Millennium* (eds J.C. Simon, C.A. Dedryver, C. Rispe & M. Hullé), pp. 45–49. Proceedings 6th International Aphid Symposium, Rennes, France, 3–7 September 2001.
- He, F. & Legendre, P. (2002) Species diversity patterns derived from species-area models. *Ecology* **83**, 1185–1198.
- Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, MA.
- Magurran, A.E. & Henderson, P.A. (2003) Explaining the excess of rare species in natural species abundance distribution. *Nature* **422**, 714–716.
- McGill, B.J. (2003a) A test of the unified neutral theory of biodiversity. *Nature* **422**, 881–885.
- McGill, B.J. (2003b) Strong and weak tests of macroecological theory. *Oikos* **102**, 679–685.
- McGuinness, K.A. (1984) Equations and explanations in the study of species–area curves. *Biological Reviews* **59**, 423–440.
- McKane, A.J., Alonso, D. & Solé, R. (2000) A mean field stochastic theory for species-rich assembled communities. *Physical Review E* **62**, 8466–8484.
- McKane, A.J., Alonso, D. & Solé, R. (2004) Analytic solution of Hubbell's model of local community dynamics. *Theoretical Population Biology* **65**, 67–73.
- Pielou, E.C. (1975) *Ecological Diversity*. John Wiley & Sons, New York.
- Pulliam, H.R. (1988) Sources, sinks, and population regulation. *American Naturalist* **132**, 652–661.
- Robbins, C.S., Bystrak, D. & Geissler, P.H. (1986) *The Breeding Bird Survey: Its First Fifteen Years, 1965–1979*. US Department of the Interior, Fish and Wildlife Services, Washington DC.
- Sauer, J.R., Hines, J.E. & Fallon, J. (2001) *The North American Breeding Bird Survey, Results and Analysis 1966–2000, Version 2001.2*. USGS, Patuxent Wildlife Research Center, Laurel, MD (available at ftp://pwrctftp.er.usgs.gov/mp/bbs/DataFiles/).
- Sugihara, G., Bersier, L.-F., Southwood, T.R.E., Pimm, S.L. & May, R.M. (2003) Predicted correspondence between species abundances and dendrograms of niche similarities. *Proceedings of the National Academy of Sciences of the USA* **100**, 5246–5251.
- Taylor, L.R. (1986) Synoptic dynamics, migration and the Rothamsted Insect Survey. *Journal of Animal Ecology* **55**, 1–38.
- Taylor, H.M. & Karlin, S. (1998) *An Introduction to Stochastic Modeling*, 3rd edn. Academic Press, San Diego, CA.
- Vallade, M. & Houchmandzadeh, B. (2003) Analytical solution of a neutral model of biodiversity. *Physical Review E* **68**, 061902–0061905.
- Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003) Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035–1037.
- Woiwod, I.P. & Harrington, R. (1994) Flying in the face of change – The Rothamsted Insect Survey. *Long Term Research in Agricultural and Ecological Sciences* (eds R.A. Leigh & A.E. Johnson), pp. 321–342. CABI, Wallingford.
- Zillinger, D. (2003) *CRC Standard Mathematical Tables and Formulae*. Chapman & Hall/CRC, Boca Raton, FL.

Received 14 June 2004; revised 4 September 2004; accepted 7 September 2004